

# Fuzzy Student's T-Distribution Model Based on Richer Spatial Combination

Tao Lei, *Senior Member, IEEE*, Xiaohong Jia, Dinghua Xue, Qi Wang, *Senior Member, IEEE*, Hongying Meng, *Senior Member, IEEE*, and Asoke K. Nandi, *Fellow, IEEE*

**Abstract**—Fuzzy c-means (FCM) algorithms with spatial information have been widely applied in the field of image segmentation. However, most of them suffer from two challenges. One is that introduction of fixed or adaptive single neighboring information with narrow receptive field limits contextual constraints leading to clutter segmentations. The other is that incorporation of superpixels with wide receptive field enlarges spatial coherency leading to block effects. To address these challenges, we propose fuzzy Student's t-distribution model based on richer spatial combination (FRSC) for image segmentation. In this Paper, we make two significant contributions. The first is that both narrow and wide receptive fields are integrated into the objective function of FRSC, which is convenient to mine image features and distinguish local difference. The second is that the rich spatial combination under Student's t-distribution ensures that spatial information is introduced into the updated parameters of FRSC, which is helpful in finding a balance between the noise-immunity and detail-preservation. Experimental results on synthetic and publicly available images, further demonstrate that the proposed FRSC addresses successfully the limitations of FCM algorithms with spatial information and provides better segmentation results than state-of-the-art clustering algorithms.

**Index Terms**—Fuzzy c-means (FCM), image segmentation, Student's t-distribution, rich spatial information.

## I. INTRODUCTION

CLUSTERING aims to assign a label to each element within a set, where elements with similar characteristics have the same label. As an unsupervised learning method, clustering requires the construction of an objective function and computes the optimized solution by minimizing the objective function. Currently, clustering has been widely used in

different fields such as image processing [1], pattern classification [2, 3], deep learning [4], etc. Among different clustering methods, fuzzy c-means (FCM) is one of the most popular methods due to its simplicity and efficiency [5-7].

Compared to  $k$ -means [8, 9], FCM has some clear advantages such as soft clustering, membership description, etc. As FCM can achieve data classification without the labeling process, it is often used for unsupervised image segmentation tasks in cases lacking Ground truth. At present, a large number of improved FCM algorithms have been proposed and successfully applied to image segmentation. However, FCM performs image segmentation similarly to normal data classification without considering image local spatial information, which leads to the fact that FCM only provides tolerable image segmentation for images with simple objects and backgrounds since it is sensitive to noise, brightness, image details, etc. To improve the performance of FCM on image segmentation, many variants of FCM algorithm have been proposed, we roughly divide these algorithms into three categories: FCM with spatial distance constraints [10-19], FCM with filtering [20-29], and FCM with Markov random field (MRF) [31-35].

FCM algorithms with spatial distance constraints improves image segmentation effect by incorporating image local spatial distance into the objective function of FCM. Researchers utilize different local spatial terms to correct membership of the central pixel within a window, which alleviates noise interference and improves segmentation performance. Based on this idea, a large number of improved FCM algorithms [10-19] are reported. In these algorithms, FCM\_S [10], FCM\_S1 [11], FCM\_S2 [11], FLICM [12], DSFCM\_N [13] and AWSFCM [14] adopt linear spatial weights obtained from experience. The RFLICM [15], KWFLICM [16], NWFCM [17], AD-FLICM [18] and PFLSM [19] employ non-linear spatial weights obtained by local similarity measurement of images. Although the latter [15-19] usually achieve better segmentation results than those of the former [10-14], the latter often suffers from higher time complexity since the computation of non-linear spatial weights is time-consuming.

FCM algorithms with filtering usually improves image segmentation results by employing pixel filtering and membership filtering. The pixel filtering is performed in the pre-processing stage and it reduces the time complexity to some extent. Popular algorithms include EnFCM [20], FGFCM [21], OSFCM-SNL [22], NDFCM [23], FRFCM [24], EnFK [25], etc. Although pixel filtering can improve image quality by suppressing noise, it also smooths image details that are often useful for segmentations, especially image contour details.

This work was supported in part by Natural Science Basic Research Program of Shaanxi (No. 2021JC-47), in part by the National Natural Science Foundation of China under Grant 61871259, Grant 61861024, Grant 62031021, and in part by Key Research and Development Program of Shaanxi (NO. 2021ZDLGY08-07). (*Corresponding author: Xiaohong Jia*)

T. Lei is with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China. (E-mail: leitao@sust.edu.cn)

X. Jia and D. Xue are with the School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China. (E-mail: jiaxhm@163.com, xdinghua@163.com)

Q. Wang is with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China. (E-mail: crabwq@gmail.com)

H. Meng is with the Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge, Middlesex, UB8 3PH, U.K. (E-mail: hongying.meng@brunel.ac.uk)

A. K. Nandi is with the Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge, Middlesex, UB8 3PH, U.K., and also with the College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China. (E-mail: asoke.nandi@brunel.ac.uk)

For this problem, the membership filtering can avoid the loss of the original image information by considering image structural information in the process of iterative optimization. Based on this idea, researchers proposed many improved FCM algorithms such as sFCM [26], csFCM [27], NCM [28], MFFCM [24], SRFCM [29], etc. These algorithms achieve better balance between noise insensitivity and image detail preservation than other improved FCM algorithms. However, their objective functions without embedding spatial information of images are similar to that of FCM.

FCM algorithms with MRF adopt pointwise prior probabilities based on regularized Kullback-Leibler information [30] to incorporate local constraints, which explores spatial coherency and local similarity of images, thus improves image segmentation performance. Existing algorithms include HMRF-FCM [31], Zhang's algorithm [32], MRFFCM [33], Liu's algorithm [34], KLDFCM [35], etc. The pointwise prior probabilities usually could be obtained by pixel label (hidden MRF) or pixel membership (MRF). The HMRF requires a large amount of computation, so researchers pay more attention to MRF prior probability since it requires low computational complexity. In addition, these algorithms generally employ negative log-likelihood of the Gaussian distribution to replace Euclidean distance or kernel distance. Therefore they inherit the advantages of Gaussian distribution and offer a significantly robust clustering results for complex data distribution.

According to the above analysis, it is popular to introduce spatial information into the objective function of FCM. However local spatial information only corresponds fixed-shape neighbored window. Fortunately, superpixel technique utilizes over-segmentation to remedy partly this problem. On the one hand, it reduces image redundancy and improves execution efficiency. On the other hand, it is convenient to capture image features and revise the quality of segmentation results. Inspired by these attributes of superpixels, researchers have successively proposed SFFCM [36], FDCM\_SSR [37] and AFCF [38]. These algorithms employ meaningful atomic regions with boundary adherence instead of fixed small neighboring to enhance the relationship between pixels which have similar local characteristics. Meanwhile, superpixels can promote the application of algorithms based on similarity matrix [37-39] in images. Although these new algorithms settle different subtasks, they rely too much on superpixel algorithms.

In this paper, we propose fuzzy Student's t-distribution model based on richer spatial combination (FRSC). The proposed FRSC reasonably takes advantage of richer spatial scheme to improve effectively image segmentation results. The main highlights of the paper are presented below.

- The FRSC utilizes negative log-likelihood of the Student's t-distribution with heavy-tail to measure distance, which can robustly fit data distribution and exhibit informative data description.
- The FRSC incorporates the pixel-level local spatial information using spatial distance constraints and membership filtering into objective function, which can collaboratively suppress noise and also meet the conditions of Lagrange optimization.

- The FRSC embeds region-level superpixels into pointwise prior probabilities. Thus the proposed algorithm not only keeps block attribute of superpixels, but also alleviates the dependence on superpixels.

The rest of this paper is organized into four sections. In the next section, we briefly review the classical variants of FCM and motivation. Section III describes the methodology of our proposed model in detail. Comparative experimental results from different algorithms are discussed in Section IV. Finally, conclusions and some open issues are outlined in Section V.

## II. RELATED WORK

FCM is one of the most mainstream algorithms based on knowledge-driven theories for unlabeled data, which was firstly proposed by Dunn [40] and finally modified by Bezdek [41]. The algorithm uses iterative optimization strategy to assign test data to  $c$  clusters. Its objective function is defined as

$$J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - \mu_i\|^2, \quad (1)$$

where  $\mathbf{X} = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{D \times n}$  is a test set of  $n$  objects with  $D$ -dimension,  $c$  is the number of clusters,  $u_{ij}$  represents the grade of membership of the  $j$ th pixel to the  $i$ th cluster and satisfies the conditions  $u_{ij} \geq 0$  and  $\sum_{i=1}^c u_{ij} = 1$ ,  $m$  is the weighting exponent of cluster fuzziness,  $\mu_i$  denotes the center of  $i$ th cluster and  $\|\cdot\|$  is a Euclidean distance. Utilizing the constraints of membership to minimize (1), we can get the corresponding membership and cluster centers as follows

$$u_{ij} = \frac{\|x_j - \mu_i\|^{\frac{-2}{m-1}}}{\sum_{k=1}^c \|x_j - \mu_k\|^{\frac{-2}{m-1}}}, \quad (2)$$

$$\mu_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m}. \quad (3)$$

FCM has a good robustness to noise-free data. Due to the diversity of pixel distribution, FCM cannot obtain ideal segmentation results when it is directly applied to image segmentation. To solve this problem, most algorithms incorporate local spatial information into their objective functions to improve segmentation performance, such as FCM\_S, MFFCM, HMRF-FCM, etc. Among them, FCM\_S introduces spatial constraints into distance measurement, while MFFCM adds the step of membership filtering in iterations. Both spatial distance constraints and membership filtering aim to smooth and revise noises or outliers. To maintain their individual strengths, we embed them both in proposed objective function to mine deeply local spatial information of images.

Similarly, HMRF-FCM also incorporate spatial modelling using pointwise prior probabilities into its objective function. Based on the fact that local spatial modelling cannot capture atomic structure of image. So we embed the superpixels with adaptive and irregular neighbors into pointwise prior probabilities. Hence, pointwise prior probabilities can not only exhibit region information, but also reflect the boundary structure of the image.

### A. Spatial Distance Constraints

In order to improve FCM for image segmentation, Ahmed *et al.* [10] proposed FCM\_S with spatial distance constraints by integrating the neighborhood information of image into its objective function. Compared with FCM, the objective function of FCM\_S is defined as follows

$$J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - \mu_i\|^2 + \frac{\alpha}{N_R} \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \sum_{r \in N_j} \|x_r - \mu_i\|^2, \quad (4)$$

where  $N_j$  denotes the index of the  $j$ th pixel's neighbor and  $N_R$  is corresponding cardinality,  $x_r$  represents the neighbor of  $x_j$  and the parameter  $\alpha$  is a penalty factor of the neighboring term. In (4), the second term presents the neighborhood information that helps FCM\_S to provide better segmentation results for images corrupted by noise. Inspired by FCM\_S, many improved algorithms have been proposed, such as FCM\_S1, FCM\_S2, FLICM, DSFCM\_N, etc.

In [10-14], most researchers believe that the high computational cost is caused by the addition of neighboring term. We could relieve the problem of time complexity from a new angle. By considering  $\|x_j - \mu_i\|_{j=1:n}^2$  as an intermediate variables image about  $\mu_i$  to relieve the problem of time complexity, it is easy to achieve the fusion of neighboring term in distance measure by the filtering with spatial weight. This simple scheme can effectively reduce the execution time and can be extended to similar improved algorithms [19].

### B. Membership Filtering

Since the membership directly decides the final classification results, FCM can be improved from the perspective of membership as well. Recently, Lei *et al.* [24] proposed MFFCM by utilizing the structural characteristics of membership to modify pixel classification. The filtering is usually implemented between (2) and (3), as follows

$$u_{ij} = \frac{u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}}{\sum_{k=1}^c (u_{kj} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{kr})}, \quad (5)$$

where  $d_{jr}$  is defined as spatial Euclidean distance between the  $j$ th pixel and the  $r$ th pixel, the spatial factor  $\frac{1}{d_{jr}+1}$  presents the decay of the neighboring membership  $u_{ir}$ . The MFFCM incorporates local spatial information in each iteration which can improve image segmentation accuracy. Many algorithms employ similar idea to improve segmentation performance, such as sFCM, csFCM, NCM, FRFCM, etc.

As we all know, FCM utilizes the Lagrange multiplier technique to minimize (1) with strict calculation. However, FCM based on membership filtering does not satisfy objective optimization theory. They can only get an approximate solution because this kind of algorithms do not introduce filtering strategy into their objective functions. How to design a rigorous algorithm of enhanced FCM still needs further research.

### C. Pointwise Prior Probabilities

By incorporating neighborhood information to prior probabilities, Chatzis *et al.* [31] introduced hidden MRF (HMRF)

model with spatial modeling to FCM, called HMRF-FCM by designing the pointwise prior probabilities. The objective function of HMRF-FCM is defined as

$$J = \sum_{i=1}^c \sum_{j=1}^n u_{ij} d_{ij} + \lambda \sum_{i=1}^c \sum_{j=1}^n u_{ij} \log\left(\frac{u_{ij}}{\pi_{ij}}\right), \quad (6)$$

where  $d_{ij}$  is negative log-posterior of a Gaussian distribution,  $\lambda$  denotes model's degree of fuzziness and  $\pi_{ij}$  denotes pointwise prior probabilities. Although the HMRF-FCM is able to improve the robustness of FCM on image segmentation, it requires high computational cost. Compared to the integration of HMRF into FCM, it is more simple and efficient to integrate spatial templates of membership into objective functions [32].

By analyzing main strategies of existing prior probability calculations, we find that most of them still adopt neighborhood information to reduce the influences of outliers. Since superpixels can provide better adaptive local spatial information, it is a good attempt to introduce superpixels attribute into the prior probability.

## III. METHODOLOGY

As mentioned above, the spatial structure information is vital for FCM in image segmentation. Under the guidance of these strategies, we propose fuzzy Student's t-distribution model based on richer spatial combination (FRSC). The proposed FRSC incorporates both spatial distance constraint  $d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir}$  and membership filtering  $u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}$  into the objective function of fuzzy clustering to effectively exploit neighborhood information of pixels. Simultaneously, the proposed FRSC also introduces the adaptive superpixels constraint [42] into the pointwise prior probability with region-level to discriminate homogeneous consistency of images.

### A. The Proposed Model

The FRSC designs the richer spatial combination including the spatial distance, the membership filtering and the pointwise prior probability. Inspired by the respective advantages of FCM\_S, MFFCM and HMRF-FCM, the objective function of FRSC is defined as

$$J = \sum_{i=1}^c \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir}) + \lambda \sum_{i=1}^c \sum_{j=1}^n u_{ij} \log\left(\frac{u_{ij}}{\pi_{ij}}\right), \quad (7)$$

where  $c$  is the number of clusters,  $n$  is the total number of pixels in an image,  $u_{ij}$  represents the fuzzy membership of the  $j$ th pixel belonging to the  $i$ th cluster, which satisfies the constraints  $u_{ij} \geq 0$  and  $\sum_{i=1}^c u_{ij} = 1$  and  $N_j$  stands for the neighborhood of the  $j$ th pixel excluding the  $j$ th pixel. Here  $d_{jr}$  is defined as spatial Euclidean distance between  $j$ th pixel and the  $r$ th pixel,  $d_{ij}$  denotes the distance function using the negative log-likelihood of Student's t-distribution and  $N_R$  is the cardinality of  $N_j$ . The parameter  $\lambda$  controls degree of fuzziness of the fuzzy membership values,  $\pi_{ij}$  is the pointwise

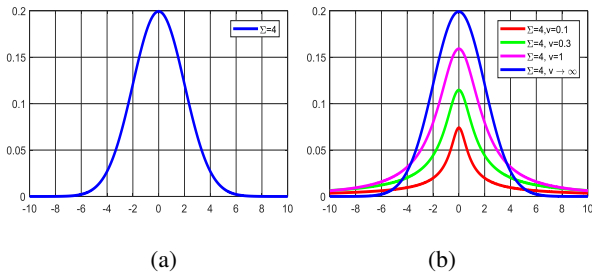


Fig. 1. Comparison between Gaussian distribution and Student's t-distribution. (a) A Gaussian distribution (the mean  $\mu = 0$  and covariance  $\Sigma = 4$ ). (b) Student's t-distribution for various values of  $v$  (the mean  $\mu = 0$  and covariance  $\Sigma = 4$ ).

prior probabilities of the  $i$ th model state with regard to the  $j$ th observation. Similar to  $u_{ij}$ , it also satisfies the constraints  $\pi_{ij} \geq 0$  and  $\sum_{i=1}^c \pi_{ij} = 1$ .

In (7), distance function is written as follows

$$d_{ij} = -\log t(x_j | \mu_i, \Sigma_i, v_i), \quad (8)$$

where  $t(x_j | \mu_i, \Sigma_i, v_i)$  denotes the probability density function of Student's t-distribution that has a longer tail and only one more parameter than the Gaussian distribution. The  $t(x_j | \mu_i, \Sigma_i, v_i)$  has its own mean vector (cluster center)  $\mu_i$ , covariance matrix  $\Sigma_i$  and degree of freedom  $v_i$ . The Student's t-distribution  $t(x_j | \mu_i, \Sigma_i, v_i)$  is defined as

$$t(x_j | \mu_i, \Sigma_i, v_i) = \frac{\Gamma(\frac{v_i + D}{2})}{\Gamma(\frac{v_i}{2})} \frac{|\Sigma_i|^{-\frac{1}{2}}}{(\pi v_i)^{\frac{D}{2}}} \left[ 1 + \frac{(x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i)}{v_i} \right]^{-\frac{v_i + D}{2}}, \quad (9)$$

where  $\Gamma(\cdot)$  represents the Gamma function,  $D$  is the dimensionality of test data  $x_j$ , and  $|\Sigma_i|$  denotes the determinant of  $\Sigma_i$ . Fig. 1 shows the possible advantages of Student's distributions.

Since Student's distributions has one more parameter (degree of freedom) compared to Gaussian distribution, it has more flexibility for data fitting. Simultaneously, if  $v \rightarrow \infty$ , the Student's t-distribution tends to a Gaussian distribution as shown in Fig. 1. From the results illustrated in Figs. 1(a)-(b), the Student's t-distribution can provide a longer tailed alternative by tuning the value of  $v$  than Gaussian distribution. Hence, Student's t-distribution is a more powerful and flexible approach for probabilistic data clustering compared to Gaussian distribution.

Furthermore, if  $v = 1$ , the Student's t-distribution is simplified to the Cauchy distribution [43]. So we can conclude that Cauchy distribution and Gaussian distribution are special cases of Student's t-distribution. The Laplace distribution uses absolute operation that is not conducive to partial derivative calculation, which limits its superiority compared to the Gaussian distribution. Through the above analysis, the Student's t-distribution shows clear superiority over other distributions.

## B. Parameters Estimation

For the FRSC, a crucial task is to obtain the optimal solution  $\Theta = \{u_{ij}, \mu_i, \Sigma_i, v_i\}$  by minimizing objective function (7). We embed the membership constraints into objective function using the Lagrange multiplier  $\gamma_j$  as follows

$$\begin{aligned} \tilde{J} = & \sum_{i=1}^c \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr} + 1} u_{ir}) (d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir}) \\ & + \lambda \sum_{i=1}^c \sum_{j=1}^n u_{ij} \log(\frac{u_{ij}}{\pi_{ij}}) + \sum_{j=1}^n \gamma_j (\sum_{i=1}^c u_{ij} - 1). \end{aligned} \quad (10)$$

To minimize the objective function, we consider the derivation of (10) on the membership. From **Appendix A**, the estimate of  $u_{ij}$  at the  $(b+1)$ th step

$$u_{ij} = \frac{\pi_{ij} \exp(-\frac{\alpha}{\lambda} (d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir}))}{\sum_{k=1}^c \pi_{kj} \exp(-\frac{\alpha}{\lambda} (d_{kj} + \sum_{r \in N_j} \frac{1}{N_R} d_{kr}))}. \quad (11)$$

According to **Appendix A**,  $\alpha = (1 + \sum_{r \in N_j} \frac{1}{d_{jr} + 1})$  is a constant for neighboring window with fixed shape. However, there is no closed form solution for  $\{\mu_i, \pi_{ij}, \Sigma_i, v_i\}$  with negative log-likelihood under Student's t-distribution. To address this problem, the Student's t-distribution [44, 45] is divided into a Gaussian distribution with mean  $\mu_i$  and covariance matrix  $\Sigma_i/t_{ij}$  and a Gamma distribution with shape  $v_i/2$  and scale  $2/v_i$ , i.e.,

$$t(x_j | \mu_i, \Sigma_i, v_i) \sim \Phi(x_j | \mu_i, \Sigma_i/t_{ij}) \Psi(t_{ij} | v_i/2, 2/v_i), \quad (12)$$

where  $t_{ij}$  is a latent precision scalar,  $\Phi(x_j | \mu_i, \Sigma_i/t_{ij})$  can be rewritten as

$$\Phi(x_j | \mu_i, \Sigma_i/t_{ij}) = \frac{\exp(-\frac{1}{2} t_{ij} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i))}{(2\pi/t_{ij})^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}}, \quad (13)$$

and  $\Psi(t_{ij} | v_i/2, 2/v_i)$  can be computed as

$$\Psi(t_{ij} | v_i/2, 2/v_i) = \frac{(t_{ij})^{\frac{v_i}{2}-1} \exp(-t_{ij}/(2/v_i))}{(2/v_i)^{v_i/2} \Gamma(v_i/2)}. \quad (14)$$

According to [46, 47], the equation (8) with negative log-likelihood can be written again as

$$\begin{aligned} d_{ij} = & \frac{D}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma_i| - \frac{D}{2} E_{\Theta}(\log t_{ij}) \\ & + \frac{1}{2} E_{\Theta}(t_{ij}) (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) + \log \Gamma(\frac{v_i}{2}) \\ & - \frac{v_i}{2} \log(\frac{v_i}{2}) - (\frac{v_i}{2} - 1) E_{\Theta}(\log t_{ij}) + \frac{v_i}{2} E_{\Theta}(t_{ij}), \end{aligned} \quad (15)$$

where  $E_{\Theta}(t_{ij})$  and  $E_{\Theta}(\log t_{ij})$  are denoted by

$$E_{\Theta}(t_{ij}) = t_{ij}^{(b)} = \frac{v_i^{(b)} + D}{v_i^{(b)} + (x_j - \mu_i^{(b)})^T (\Sigma_i^{-1})^{(b)} (x_j - \mu_i^{(b)})}, \quad (16)$$

and

$$E_{\Theta}(\log t_{ij}) = \log t_{ij}^{(b)} - \log(\frac{v_i^{(b)} + D}{2}) + \psi(\frac{v_i^{(b)} + D}{2}), \quad (17)$$

where  $\psi(s) = \frac{\partial \log(\Gamma(s))}{\partial s}$  is the di-gamma function [44, 45]. Observing (16) and (17), we find that the current  $(b+1)$ th step

estimation of the algorithm requires corresponding to variables at the previous  $(b)th$  step. By substituting (16, 17) to (15), the estimate of  $d_{ij}$  at the  $(b+1)th$  step can be obtained

$$d_{ij} = \frac{D}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma_i| + \log \Gamma\left(\frac{v_i}{2}\right) + \frac{v_i}{2} t_{ij}^{(b)} + \frac{1}{2} t_{ij}^{(b)} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) - \frac{v_i}{2} \log\left(\frac{v_i}{2}\right) - \left(\frac{v_i + D}{2} - 1\right) \left(\log t_{ij}^{(b)} - \log\left(\frac{v_i^{(b)} + D}{2}\right) + \psi\left(\frac{v_i^{(b)} + D}{2}\right)\right), \quad (18)$$

By combining (7) and (18), we can obtain (19).

From **Appendix B**, the clustering centers  $\mu_i$  at  $(b+1)th$  step are obtained

$$\mu_i = \frac{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} x_j + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} x_r)}{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)})}. \quad (20)$$

Similarly, from **Appendix C**, the covariance matrix  $\Sigma_i$  at  $(b+1)th$  step is obtained

$$\Sigma_i = \frac{\sum_{j=1}^n \left[ (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} (x_j - \mu_i)^T (x_j - \mu_i) + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} (x_r - \mu_i)^T (x_r - \mu_i)) \right]}{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (1 + \sum_{r \in N_j} \frac{1}{N_R})}. \quad (21)$$

In addition, from **Appendix D**, the degree of freedom  $v_i$  at  $(b+1)th$  step can be calculated as follows:

$$\psi\left(\frac{v_i}{2}\right) + \frac{\sum_{j=1}^n \left[ (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) ((t_{ij}^{(b)} - \log t_{ij}^{(b)}) + \sum_{r \in N_j} \frac{1}{N_R} (t_{ir}^{(b)} - \log t_{ir}^{(b)})) \right]}{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (1 + \sum_{r \in N_j} \frac{1}{N_R})} - \log\left(\frac{v_i}{2}\right) - 1 + \log\left(\frac{v_i^{(b)} + D}{2}\right) - \psi\left(\frac{v_i^{(b)} + D}{2}\right) = 0. \quad (22)$$

The spatial distance constraints and membership filtering are integrated into the objective function of the FRSC, which guarantees spatial structure information being implemented in all variables, as shown in formula (11) and (20)-(22). During the convergence process of the FRSC, the small window avoids over-smoothing and keeps the richer original characteristics, which can mine deeper hidden feature to guide correct image segmentation.

### C. Prior Probabilities

To affect result at  $(b+1)th$  iteration of the proposed algorithm, it is easy to design pointwise prior probabilities according to ideas in [31, 32], i.e.,

$$\pi'_{ij} = \frac{u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}}{\sum_{k=1}^c (u_{kj} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{kr})}, \quad (23)$$

In (23), the  $\pi'_{ij}$  only considers local information within narrow receptive field, which limits the capability of pointwise prior probabilities. To solve the problem, superpixels is often employed since it provides meaningful atomic regions with accurate boundaries, and present homogeneous features within a wide receptive field as shown in Fig. 2.

According to Fig. 2, we can find that traditional neighboring information can perceive only regular narrow filed, which ignores the really local-spatial structure of image. In order to overcome this shortcoming, we can use superpixel region generated by AMR-WT to capture wide receptive filed.

Inspired by this, we employ the property of superpixels to modify pointwise prior probabilities, and thus get a new pointwise prior probability,

$$\pi_{ij} = \frac{\frac{1}{S_j} \sum_{l \in R_j} (u_{il} + \sum_{r \in N_l} \frac{1}{d_{lr}+1} u_{il})}{\sum_{k=1}^c \frac{1}{S_j} \sum_{l \in R_j} (u_{kl} + \sum_{r \in N_l} \frac{1}{d_{lr}+1} u_{kl})}, \quad (24)$$

where  $S_j$  represents the pixel number in region  $R_j$  including the  $j$ th pixel of an image, the adaptive region  $R_j$  is provided by the AMR-WT [43]. The  $u_{il}$  stands for fuzzy membership of the  $l$ th pixel within the region  $R_j$  with respect to the  $i$ th cluster. Since adaptive regions provide better spatial relationship than local neighborhood information within fixed windows, the  $\pi_{ij}$  can help (11) to improve image segmentation effect.

### D. Algorithm Framework of the FRSC

According to the analysis above, we get richer spatial combination. The spatial constraints with small window can lightly smooth outliers. Meanwhile, superpixels with blocks can revise boundaries of an image. The richer spatial combination not only retains advantages of local structure, but also exhibits the attributes of superpixels.

To illustrate clearly our algorithm, we summarize these parameters  $\Theta = \{u_{ij}, \mu_i, \pi_{ij}, \Sigma_i, v_i\}$  updates of our algorithm as follows

- (1) Set the number of clusters  $c$ , set the convergence threshold  $\eta$ , and set the maximal iteration number  $B$ .
- (2) Before iteration, get a superpixels result from AMR-WT and select the neighboring window. In addition, initialize the parameters, including the cluster centers  $\mu_i$ , the pointwise prior probabilities  $\pi_{ij}$ , the covariance matrix  $\Sigma_i$  and the degree of freedom  $v_i$ .
- (3) Set the loop counter  $b = 0$ .
- (4) Update the membership matrix  $u_{ij}$  using (11);
- (5) Update the latent variable  $t_{ij}$  using (16);
- (6) Update the cluster centers  $\mu_i$  using (20);
- (7) Update the covariance matrix  $\Sigma_i$  using (21);
- (8) Update the pointwise prior probabilities  $\pi_{ij}$  using (24);
- (9) Update the degree of freedom  $v_i$  using (22);
- (10) If  $\max |J^{(b)} - J^{(b-1)}| \leq \eta$  or  $b \geq B$ , stop; otherwise, update  $b = b + 1$ , and go to step 4.

When the proposed algorithm meets the convergence condition, we can obtain the optimal membership matrix, then utilize a de-fuzzification process by maximum membership to convert membership matrix to label matrix. Since  $u_{ij}$  is

$$\begin{aligned}
J = & \sum_{i=1}^c \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (\frac{D}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma_i| + \log \Gamma(\frac{v_i}{2}) - \frac{v_i}{2} \log(\frac{v_i}{2}) + \frac{1}{2} t_{ij}^{(b)} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) \\
& + \frac{v_i}{2} t_{ij}^{(b)} - (\frac{v_i + D}{2} - 1) (\log t_{ij}^{(b)} - \log(\frac{v_i^{(b)} + D}{2}) + \psi(\frac{v_i^{(b)} + D}{2}))) + \sum_{r \in N_j} \frac{1}{N_R} (\frac{D}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma_i| + \log \Gamma(\frac{v_i}{2}) \\
& - \frac{v_i}{2} \log(\frac{v_i}{2}) + \frac{1}{2} t_{ir}^{(b)} (x_r - \mu_i)^T \Sigma_i^{-1} (x_r - \mu_i) + \frac{v_i}{2} t_{ir}^{(b)} - (\frac{v_i + D}{2} - 1) (\log t_{ir}^{(b)} - \log(\frac{v_i^{(b)} + D}{2}) + \psi(\frac{v_i^{(b)} + D}{2}))).
\end{aligned} \tag{19}$$

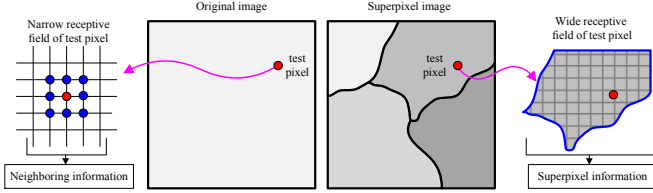


Fig. 2. Illustration of different between narrow receptive field and wide receptive field for test pixel.

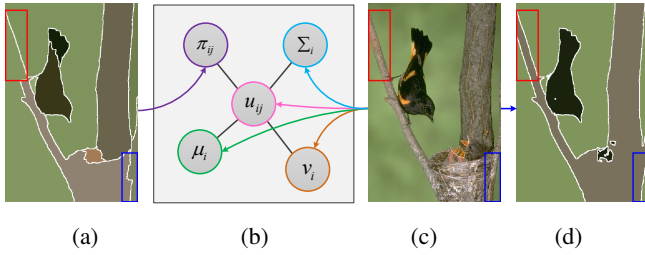


Fig. 3. Segmentation process of the FRSC. (a) AMR-WT. (b) Clustering process of the FRSC. (c) Original image. (d) Segmentation result.

replaced by  $u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}$ , we assign pixels to the class  $L$  with the spatial highest membership value, as follows

$$L_j = \arg_{\{L\}} \{ \max \{ u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir} \} \}. \tag{25}$$

According to  $L_j$ , the fuzzy membership matrix is converted to label image.

To illustrate further the strengths of the FRSC, Fig. 3 shows the process of segmentation. In Fig. 3(a), there are two serious region-level defects marked by red and blue boxes. The proposed FRSC utilizes pixel-level neighboring substructures of original image to avoid segmentation errors as shown in Fig. 3(d). But Fig. 3(d) also maintains some attributes of superpixels. The FRSC combines their respective advantages, called region-level blocks and pixel-level neighboring substructures, to improve final segmentation effect.

#### IV. EXPERIMENTS

In this Section, we conducted experiments on synthetic images, real color images from different datasets that include the Berkeley Segmentation Dataset and Benchmark (BSDS) [48], and the Microsoft Research Cambridge (MSRC) [49]. To demonstrate the superiority of proposed algorithm, we

TABLE I. Dice index of FRSC with different parameters, where  $\alpha$  is a constant in (11).

$\lambda$	$w \times w$	$3 \times 3$	$5 \times 5$	$7 \times 7$	$9 \times 9$
$\alpha$		0.9940	0.9938	0.9935	0.9931
$2\alpha$		0.9965	0.9964	0.9959	0.9954
$3\alpha$		0.9891	0.9889	0.9886	0.9881
$4\alpha$		0.9769	0.9764	0.9760	0.9755

also evaluated FCM\_S [10], HMRF-FCM [31], FLICM [12], KWFLICM [16], Liu's algorithm [34], FRFCM [24], MF-FCM [24], DSFCM\_N [13], SFFCM [36] for comparison. To make all test algorithms obtain better segmentation effect, we adopted the CIELab color space. Our experiments are executed with MATLAB R2019, and are conducted on a DELL desktop with Intel(R) Core (TM) CPU, i7-6700, 3.4GHz, 16GB RAM.

##### A. Parameter setting

Before accomplishing all numerical experiments, we firstly depict parameter setting of all test algorithms. For fuzzy clustering algorithms, there are three essential parameters, namely the weighting exponent, the convergence threshold, and the maximal number of iterations. In our experiments, the three parameters are set as 2,  $10^{-5}$  and 100, respectively. In addition, a local window of size  $3 \times 3$  is fairly adopted for all algorithms considering spatial information.

Except for three essential parameters and the window size, there are no more parameters used for FLICM, KWFLICM and MFFCM. Nevertheless, the penalty factor is often used to control the effect of neighboring term in the FCM\_S, and its value is set to 3. For the HMRF-FCM and Liu's algorithm, the degree of fuzziness is usually neglected since its value is 1. Meanwhile the Liu's algorithm employs initial over-segmentation provided by the mean-shift [50] that is implemented with the spatial bandwidth  $h_s = 10$ , the range bandwidth  $h_r = 10$ , and the minimum region area  $h_k = 100$ . According to the DSFCM\_N, the regularization parameter is chosen as one quarter of the standard deviation of each channel in image. The structural element used for FRFCM follows the original paper. In the SFFCM, the minimal structuring element is a disk and its radius is set to 2.

Except for the convergence threshold and the maximal number of iterations, the objective function of proposed FRSC still includes two main parameters: the size of local window



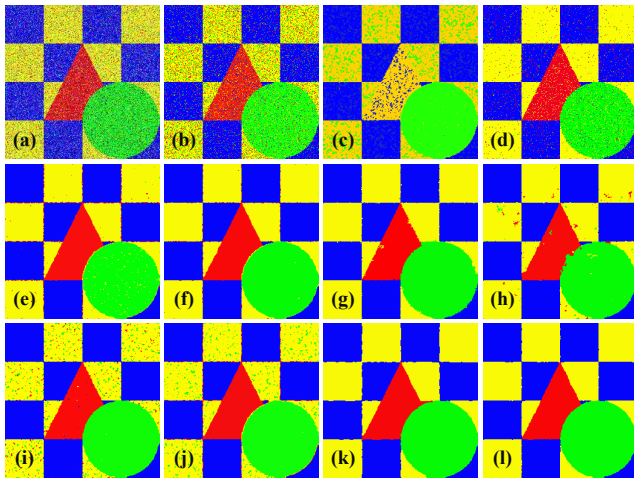


Fig. 4. The comparison of segmentation results using different algorithms on the first synthetic image with four-class. (a) Noisy image (SP noise with density is 10% and G noise with zero mean and 10% variance). (b) FCM. (c) FCM\_S. (d) HMRF-FCM. (e) FLICM. (f) KWFLICM. (g) Liu's algorithm. (h) FRFCM. (i) MFFCM. (j) DSFCM\_N. (k) SFFCM. (l) FRSC.

$w$  and the parameter  $\lambda$  that controls degree of fuzziness. To explain clearly the selection for  $w$  and  $\lambda$ , we exhibit a series of experiments as shown in Table I. Fig. 4(a) is tested after Salt and Pepper noise (SP), Gaussian noise (G), and Rician noise (R) have been corrupted, while their noise levels are 15%. Table I shows that the Dice index of FRSC depends on the two parameters. If  $\lambda$  is fixed, as the window size increases, Dice index will get worse and the time complexity will get higher. So we should choose a small window for FRSC. If the window size is fixed, we find that when  $\lambda = 2\alpha$ , the performance of FRSC is the best. According to the comparative experiment in table I, we use  $3 \times 3$  window and  $\lambda = 2\alpha$ . Simultaneously, AMR-WT is considered as a pre-segmentation with the minimal structuring element of 2 to impact on pointwise prior probabilities.

### B. Synthetic Images

To compare the robustness of different algorithms for noise and mosaic, we choose two synthetic images of size  $256 \times 256$ , as shown in Fig. 4(a) with mixture noise and Fig. 5(a) with texture mosaic. The mixture noise is the composition of Salt and Pepper noise (SP) and Gaussian noise (G). These textures are chosen from the MIT Media Lab VisTex <https://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>. Segmentation results of all mentioned algorithms are shown in Figs. 4-5(b-l).

Fig. 4(b) shows that the FCM is sensitive to noise because of ignoring spatial information of images. FCM\_S fails to segment correct class as shown in Fig. 4(c). Figs. 4(d, i, g) indicate that HMRF-FCM, MFFCM and DSFCM\_N can suppress noise to some extent, but these segmentation results still contain a lot of isolated regions caused by noise. In contrast with these clustering algorithms, both FLICM and FRFCM obtain better segmentation results as shown in Figs.

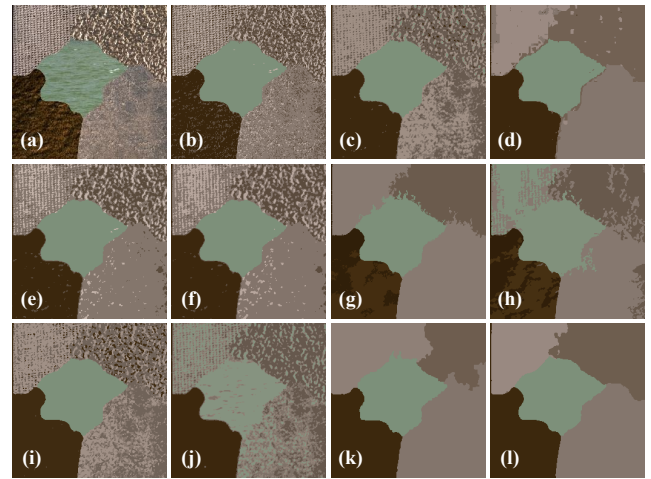


Fig. 5. The comparison of segmentation results using different algorithms on the second synthetic image with five-class. (a) Texture image (these textures are from the MIT Media Lab VisTex). (b) FCM. (c) FCM\_S. (d) HMRF-FCM. (e) FLICM. (f) KWFLICM. (g) Liu's algorithm. (h) FRFCM. (i) MFFCM. (j) DSFCM\_N. (k) SFFCM. (l) FRSC.

4(e, h). As the KWFLICM employs tradeoff weighted fuzzy factor and kernel metric, it provides favorable segmentation as shown in Fig. 4(f). Compared to Fig. 3(f), Figs. 3(g, k) show better visual effect since both Liu's algorithm and the SFFCM employ superpixels algorithms to obtain excellently adaptive spatial neighborhood information. However, Figs. 4(g, k) still contain some misclassified pixels. Compared to other algorithms, the proposed FRSC employs richer spatial combination information and thus achieves the best segmentation result as shown in Fig. 4(l).

For real synthetic images, texture information plays an important role for the final segmentation effect. Fig. 5 shows the comparison of segmentation results using different algorithms on the synthetic image with real texture information. Except HMRF-FCM, Liu's algorithm and SFFCM, other comparative algorithms are less immune to textures mosaic leading to a large amount of clutter as shown in Figs. 5(b, c, e, f, h, i, j). The main reason is that these algorithms only employ one kind of spatial information. Fig. 5(d) shows the segmentation of HMRF-FCM, which contains a few spots in consistent regions. Although Liu's algorithm avoids the clutter to some extent as shown in Fig. 5(g), it confuses classes due to little difference of texture mosaic. Focusing on Fig. 5(k) generated by the SFFCM, we find that the segmentation result contains clear boundaries and shows good visual effect. Furthermore, the segmentation result shown in Fig. 5(l) is superior to the Fig. 5(k) since the FRSC utilizes richer spatial combination information than comparative algorithms.

To evaluate quantitatively the performance of all test algorithms, we add different levels of noise on the first synthetic image and use segmentation accuracy (SA), Dice index (DI), partition coefficient (PC) and partition entropy (PE) to evaluate the performance of different algorithms. Figs. 6-7 show the comparison of performance metrics. For the second synthetic

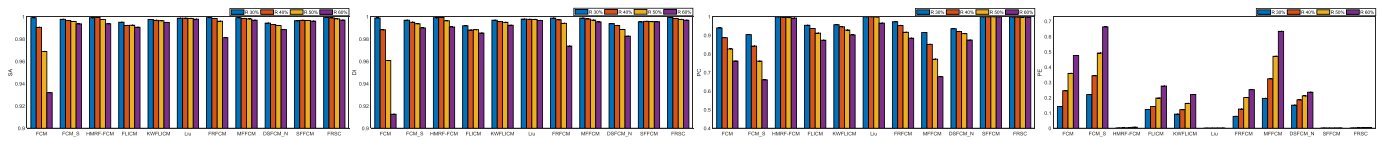


Fig. 6. Performance comparison of different algorithms on the first synthetic image corrupted by Rician noise with different levels.

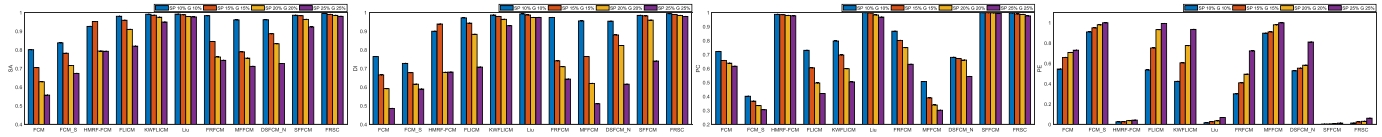


Fig. 7. Performance comparison of different algorithms on the first synthetic image corrupted by mixture noise (Salt &amp; Pepper noise and Gaussian noise) with different levels.

TABLE II. Segmentation performances of different algorithms on the second synthetic image with texture mosaic. The best values are in bold.

Indices	FCM	FCM_S	HMRF-FCM	FLICM	KWFLICM	Liu's algorithm	FRFCM	MFFCM	DSFCM_S	SFFCM	FRSC
S $\uparrow$	0.5501	0.5134	0.9058	0.6006	0.6247	0.6814	0.5896	0.4893	0.4602	0.8495	<b>0.9594</b>
SA $\uparrow$	0.7098	0.6785	0.9506	0.7505	0.7690	0.8296	0.7911	0.6571	0.6303	0.9186	<b>0.9793</b>
Dice $\uparrow$	0.7007	0.6895	0.9481	0.7111	0.7520	0.7942	0.7518	0.6633	0.6230	0.9166	<b>0.9777</b>
NMI $\uparrow$	0.5202	0.5535	0.8624	0.6344	0.6477	0.7745	0.6567	0.5363	0.4553	0.8183	<b>0.9313</b>
F-score $\uparrow$	0.7022	0.7101	0.9490	0.7493	0.7741	0.8351	0.7742	0.6931	0.6337	0.9224	<b>0.9779</b>
PC $\uparrow$	0.6532	0.4079	0.9742	0.6351	0.6673	0.9850	0.7502	0.4404	0.4698	0.9867	<b>0.9822</b>
PE $\downarrow$	0.7099	0.9103	<b>0.0111</b>	0.7432	0.6862	0.0280	0.5129	0.9002	0.9714	0.0352	0.0220

image with texture mosaic, we add three popular indicators, i.e., quantitative index score (S), normalized mutual information (NMI) and F-score, to evaluate image segmentation effect provided by different algorithms. Except for PE, a larger value of indicators corresponds to better segmentation effect. Table II shows the performance metrics of different algorithms.

As illustrated in Fig. 6, HMRF-FCM, Liu's algorithm, SFFCM and FRSC perform well on the first synthetic image corrupted by Rician noise of different levels. Especially, the four algorithms almost provide extreme values of PC and PE due to their employment of negative log-likelihood function and superpixel. A similar situation also appears in Fig. 7. By comparing Figs. 6-7, we can infer that segmentation results of FRSC are better than other comparative algorithms, and its membership values are tending closer 1.

Different from Figs. 6-7, Table II shows seven indicators of all algorithms for synthetic image with texture mosaic. Except for PC and PE, the change trends of other indicators are relatively close. The satisfaction of FCM\_S, MFFCM, and DSFCM\_N is even lower than that of FCM as shown in Table II. FLICM, KWFLICM and FRFCM also provide poor indicators. However, the HMRF-FCM and SFFCM perform well for different indicators. It is worth noting that the Liu's algorithm is not stable. Consistent with the visual effect of Fig. 5(l), the proposed FRSC outperforms other comparative algorithms due to the use of richer spatial combination information.

### C. Benchmarks

In Section IV.B, we conducted experiments on two synthetic images, and experimental results demonstrate that the proposed FRSC outperforms comparative algorithms. To demon-

strate further the superiority of the proposed algorithm, we conducted experiments on two popular Benchmarks BSDS500 and MSRC. The BSDS500 includes 500 natural images with size of either  $481 \times 321$  or  $321 \times 481$ , and each image corresponds to 4-9 manually labeled ground truths. The MSRC collects 591 images with size of either  $320 \times 213$  or  $213 \times 320$  and covers 23 object classes. There is only one ground truth for each image in MSRC.

Fig. 8 visually shows segmentation results of comparative algorithms on BSDS500. As can be seen from Fig. 8, FCM without spatial information of image is sensitive to clutter in color image. Segmentation results obtained by FCM\_S, HMRF-FCM, FLICM, KWFLICM, FRFCM and MFFCM include lots of isolated small areas due to the employment of a small local window. DSFCM\_N also adopts a similar strategy, but it only works for part of image, e.g, the first test image in Fig. 8. Liu's algorithm and SFFCM generate better segmentation results than aforementioned algorithms, which are attributed to the incorporation of superpixels. These methods utilize spatial information and superpixels algorithm to relieve the interferences from complex background and low-intensity contrast. For the proposed FRSC, the richer spatial combination explores the advantages of spatial information and superpixels, which deeply mines the potential information of image to improve the clustering results as shown in the last row of Fig. 8.

In Fig. 9, some test images from the MSRC shows simpler background and texture than Fig. 8. Similarly, the proposed FRSC achieves the best segmentation results of all the compared algorithms. Especially the seventh row of Fig. 9 presents incredible contrast, which verifies that the FRSC with richer





Fig. 8. Comparison of segmentation results using different algorithms on color images from the BSDS500.

spatial combination can obtain segmentation objects with accurate boundaries. From Figs. 8-9, we also discover that comparative algorithms are unstable for different scenes. On the contrary, the FRSC employing richer spatial combination not only retains local characteristics of images, but also avoids block effect caused by superpixels, which helps the proposed algorithm to generate stable and good segmentation results.

To evaluate further the segmentation performance of all testing algorithms on the BSDS500 and the MSRC, we adopt five performance measures [36]. The probabilistic rand index (PRI), the coving (CV), the variation of information (VI), the global consistency error (GCE), and the boundary displacement error (BDE) are used as metric criterion. The PRI measures the similarity between segmentation result and ground truth. The CV is an overlap measure of two segmentation results. The VI focuses on the distance of two partitions in terms of their entropy. The GCE estimates the deviation between two segmentations. The BDE assesses boundary error

of two segmentations results. When the segmentation result is close to the ground truth, then the value of PRI and CV will be large while the value of VI, GCE and BDE will be small.

According to the size of testing images, we set the number of clusters from 2 to 6 for an image in BSDS500, while we set the number of clusters from 2 to 4 for an image in MSRC. We select a group of data corresponding to the best PRI for each image. Tables III-IV show the comparison of performance metrics on two datasets, where we utilize mean and standard deviation (std) to estimate stability of algorithms. Through the qualitative and quantitative evaluations, we note that DSFCM\_N obtains lower values of metrics since the sparse deviation with local spatial information causes serious error with a higher probability. In addition to the SFFCM, other comparative algorithms obtain similar values of five metrics. Compared with SFFCM, the proposed FRSC achieves overall the most excellent mean values of PRI, CV, VI, GCE and BDE. Meanwhile, the standard deviations of FRSC

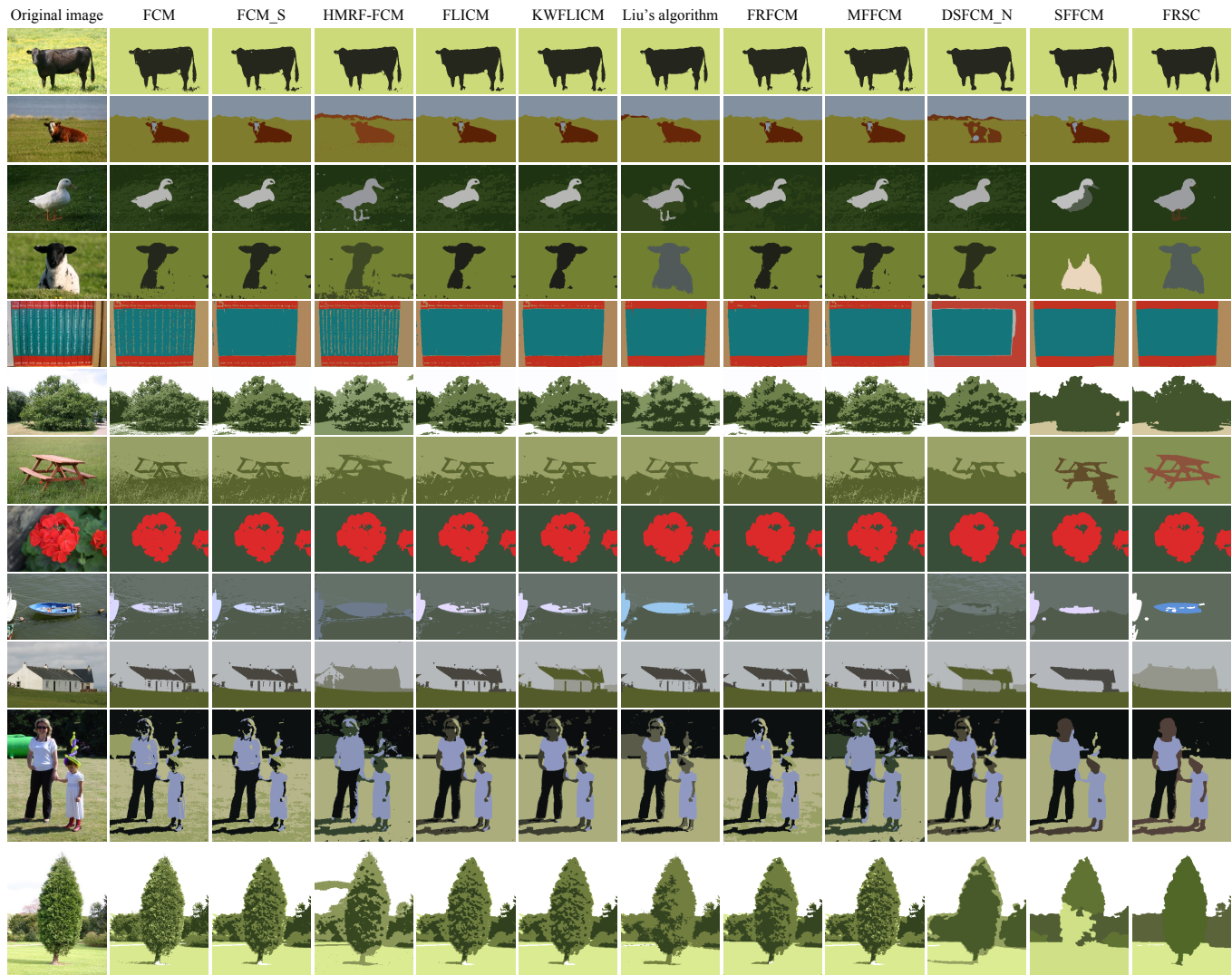


Fig. 9. Comparison results of different algorithms on color images from the MSRC.

TABLE III. Average metrics (“ $\pm$ ”std) of comparative algorithms on the BSDS500 dataset.

Algorithms	PRI $\uparrow$	CV $\uparrow$	VI $\downarrow$	GCE $\downarrow$	BDE $\downarrow$
FCM	0.74 $\pm$ 0.09	0.43 $\pm$ 0.18	2.88 $\pm$ 1.04	0.40 $\pm$ 0.16	13.48 $\pm$ 6.51
FCM_S	0.75 $\pm$ 0.09	0.43 $\pm$ 0.18	2.83 $\pm$ 1.04	0.40 $\pm$ 0.16	13.37 $\pm$ 6.85
HMRF-FCM	0.74 $\pm$ 0.09	0.43 $\pm$ 0.17	2.77 $\pm$ 1.02	0.40 $\pm$ 0.16	13.22 $\pm$ 6.90
FLICM	0.74 $\pm$ 0.09	0.43 $\pm$ 0.17	2.77 $\pm$ 1.02	0.40 $\pm$ 0.16	13.22 $\pm$ 6.90
KWFLICM	0.74 $\pm$ 0.09	0.44 $\pm$ 0.17	2.83 $\pm$ 1.01	0.40 $\pm$ 0.16	13.40 $\pm$ 6.48
Liu's algorithm	0.76 $\pm$ 0.09	0.47 $\pm$ 0.17	2.58 $\pm$ 0.99	0.36 $\pm$ 0.16	12.31 $\pm$ 6.79
FRFCM	0.76 $\pm$ 0.09	0.45 $\pm$ 0.17	2.67 $\pm$ 0.98	0.37 $\pm$ 0.16	12.35 $\pm$ 6.85
MFFCM	0.75 $\pm$ 0.09	0.44 $\pm$ 0.17	2.78 $\pm$ 1.00	0.39 $\pm$ 0.16	13.40 $\pm$ 6.73
DSFCM_N	0.74 $\pm$ 0.10	0.42 $\pm$ 0.17	2.90 $\pm$ 1.02	0.41 $\pm$ 0.17	13.83 $\pm$ 7.42
SFFCM	0.78 $\pm$ 0.10	0.55 $\pm$ 0.16	2.06 $\pm$ 0.92	0.26 $\pm$ 0.15	12.80 $\pm$ 9.61
FRSC	<b>0.81<math>\pm</math>0.08</b>	<b>0.57<math>\pm</math>0.16</b>	<b>1.98<math>\pm</math>0.79</b>	<b>0.24<math>\pm</math>0.13</b>	<b>11.20<math>\pm</math>6.74</b>

TABLE IV. Average metrics (“ $\pm$ ”std) of comparative algorithms on the MSRC dataset.

Algorithms	PRI $\uparrow$	CV $\uparrow$	VI $\downarrow$	GCE $\downarrow$	BDE $\downarrow$
FCM	0.70 $\pm$ 0.10	0.55 $\pm$ 0.17	1.93 $\pm$ 0.75	0.32 $\pm$ 0.15	12.67 $\pm$ 7.22
FCM_S	0.70 $\pm$ 0.11	0.56 $\pm$ 0.17	1.85 $\pm$ 0.74	0.31 $\pm$ 0.15	12.51 $\pm$ 7.38
HMRF-FCM	0.70 $\pm$ 0.11	0.56 $\pm$ 0.16	1.84 $\pm$ 0.62	0.31 $\pm$ 0.14	12.38 $\pm$ 7.86
FLICM	0.72 $\pm$ 0.10	0.59 $\pm$ 0.16	1.73 $\pm$ 0.65	0.28 $\pm$ 0.14	12.29 $\pm$ 7.68
KWFLICM	0.69 $\pm$ 0.10	0.55 $\pm$ 0.16	1.93 $\pm$ 0.64	0.32 $\pm$ 0.13	12.67 $\pm$ 7.58
Liu's algorithm	0.71 $\pm$ 0.11	0.54 $\pm$ 0.17	1.77 $\pm$ 0.66	0.34 $\pm$ 0.14	12.43 $\pm$ 8.60
FRFCM	0.73 $\pm$ 0.10	0.62 $\pm$ 0.15	1.79 $\pm$ 0.67	0.30 $\pm$ 0.14	12.23 $\pm$ 8.18
MFFCM	0.70 $\pm$ 0.10	0.57 $\pm$ 0.17	1.82 $\pm$ 0.72	0.30 $\pm$ 0.15	12.65 $\pm$ 7.45
DSFCM_N	0.69 $\pm$ 0.11	0.54 $\pm$ 0.18	1.91 $\pm$ 0.75	0.32 $\pm$ 0.17	12.70 $\pm$ 7.96
SFFCM	0.73 $\pm$ 0.11	0.62 $\pm$ 0.18	1.58 $\pm$ 0.64	0.25 $\pm$ 0.14	12.49 $\pm$ 9.01
FRSC	<b>0.76<math>\pm</math>0.10</b>	<b>0.65<math>\pm</math>0.16</b>	<b>1.45<math>\pm</math>0.59</b>	<b>0.22<math>\pm</math>0.13</b>	<b>10.80<math>\pm</math>7.82</b>

are lightly better than most comparative algorithms, which demonstrates the stability of the FRSC. Figs. 8-9 and Tables III-IV demonstrate further the effectiveness of the proposed FRSC for real image segmentation.

Since statistical significance of an algorithm can exhibit that the differences observed in experiments are real and not due to chance, we calculate the  $p$  value by conducting  $t$ -test to describe the probability of differences between two groups.

Then we perform statistical significance test of the proposed FRSC versus other comparison algorithms on BSDS500. We find that the  $p$  values of the five indicators (PRI, CV, VI, GCE, BDE) are all less than 0.01. For MSRC, we can get similar results. In most sciences, if the  $p \leq 0.01$ , results are generally considered statistically significant. Through the above analysis, it can be concluded that the proposed algorithm is statistically significant.

#### D. Computational cost

The computational complexity is important for estimating the practicability of algorithms. Table V shows the comparison of computational complexity for different algorithms, where  $n$  is the number of pixels of test images,  $c$  is the number of clusters,  $b$  is the number of iteration,  $w$  is the size of local window,  $O(n \times \tau)$  is the execution cost of initial over-segmentation for Liu's algorithm,  $O(n \times t')$  is the computational complexity of AMR-WT,  $n'$  is the total number of superpixel areas and  $n' \ll n$ . Although FCM\_S, HMRF-FCM, FLICM and MFFCM employ different local neighbor information, they have the same time complexity.

As shown in Table V, FCM\_S with local spatial information requires higher computational burden than FCM in each iteration. In order to get the local coefficient of variation, KWFLICM needs longer computational time. SFFCM uses a superpixel technology, so it has lower time complexity than FRFCM. Compared to Liu's algorithm, FRSC mines more spatial information from membership and distance, but it simplifies the calculation of pointwise prior probabilities.

To verify further the above analysis, we compare the execution time of eleven algorithms for different images as shown in Table VI. With BSDS500 and MSRC, we compute the average running time of all testing images within each benchmark dataset. In addition,  $c$  is set to 2-6 for BSDS500 while  $c$  is set to 2-4 for MSRC. It can be seen from Table VI that SFFCM requires the shortest execution time, even lower than FCM, due to the fact that the number of superpixel areas is much smaller than the number of pixels. Obviously, FCM\_S, FLICM, FRFCM, MFFCM, and DSFCM\_N exhibit similar high computational efficiency. On the contrary, HMRF-FCM and KWFLICM need long execution time since hidden Markov models and local coefficient of variation are considered. The proposed FRSC is faster than Liu's algorithm but slower than most comparison algorithms; this drawback can be compensated by its very good performance as mentioned above.

#### V. CONCLUSION

In this paper, fuzzy Student's t-distribution model based on richer spatial combination (FRSC) has been proposed. Through combining spatial distance constraints, membership filtering and pointwise prior probabilities with superpixels, the proposed algorithm not only can reduce noise interference, but also can improve segmentation quality. The FRSC takes advantages of each module and balances the respective role in image segmentation. Thus, with the help of pixel-level neighboring substructures and region-level block attribute,

TABLE V. Comparison of computational complexity of different algorithms.

Algorithms	Computational complexity
FCM	$O(n \times c \times b)$
FCM_S	$O(n \times c \times w^2 \times b)$
KWFLICM	$O(n \times (w+1)^2 + n \times c \times w^2 \times b)$
Liu's algorithm	$O(n \times \tau + n \times c \times w^2 \times b)$
FRFCM	$O(n \times w^2 + n \times c \times t)$
SFFCM	$O(n \times t' + n' \times c \times t)$
FRSC	$O(n \times t' + n \times c \times (2 \times 2) \times b)$

the FRSC is able to determine details or boundaries from interference. In addition, the algorithm adopts the negative log-likelihood under Student's t-distribution as distance measure, which is more robust for outliers or noises. Therefore, the proposed algorithm can deeply mine comprehensive neighboring features and obtain much improved segmentation results. In consequence, the FRSC is obviously better than the popular comparison algorithms in different kinds of images.

There are some open issues worth discussing. Firstly, the FRSC employs fixed neighboring weight, which can reduce the execution time, but it also weakens local differences. Secondly, the FRSC lacks sparse membership which better distinguish similarities of data. How to design adaptive neighboring weight, obtain sparse membership, and deal with related problems would require more research efforts.

#### APPENDIX A

##### PROOF OF MEMBERSHIP MATRIX $u_{ij}$

In order to simplify the calculation, we create the temporary function of  $\tilde{J}_{u_{ij}}$  with respect to  $u_{ij}$  according to (10)

$$\begin{aligned} \tilde{J}_{u_{ij}} = & \sum_{i=1}^c \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr} + 1} u_{ir}) (d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir}) \\ & + \lambda \sum_{i=1}^c \sum_{j=1}^n u_{ij} \log(\frac{u_{ij}}{\pi_{ij}}) + \gamma_j \sum_{i=1}^c u_{ij}, \end{aligned} \quad (\text{A.1})$$

According to the correlation between neighborhood information in Fig. 10, we can obtain that  $\frac{\partial \sum_{i=1}^c \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr} + 1} u_{ir})}{\partial u_{ij}} = 1 + 0.5 \times 4 + \frac{1}{\sqrt{(2)+1}} \times 4$  under a  $3 \times 3$  window. On this basis, we can get

$$\frac{\partial \sum_{i=1}^c \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr} + 1} u_{ir})}{\partial u_{ij}} = 1 + \sum_{r \in N_j} \frac{1}{d_{jr} + 1} \quad (\text{A.2})$$

The derivative of A.1 with respect to  $u_{ij}$  is given as follows

$$\begin{aligned} \frac{\partial \tilde{J}_{u_{ij}}}{\partial u_{ij}} = & (1 + \sum_{r \in N_j} \frac{1}{d_{jr} + 1}) (d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir}) \\ & + \lambda \log(\frac{u_{ij}}{\pi_{ij}}) + \lambda + \gamma_j = 0, \end{aligned} \quad (\text{A.3})$$

TABLE VI. Comparison of execution time (in seconds) of eleven algorithms for different images.

Noise	FCM	FCM_S	HMRf-FCM	FLICM	KWFLICM	Liu's algorithm	FRFCM	MFFCM	DSFCM_N	SFFCM	FRSC
Fig. 4	0.8631	1.8084	22.2269	1.4434	21.7421	12.0329	0.9587	1.0666	5.7971	0.2286	9.0133
Fig. 5	1.1206	2.0347	28.0280	1.9406	30.2327	20.6112	1.2202	1.4044	7.8515	0.2776	16.9587
BSDS500	2.1880	3.7038	52.5729	3.3431	69.0186	77.6384	3.6701	2.9875	17.1021	0.8251	38.2293
MSRC	0.5209	0.8617	16.1072	1.1458	23.8346	21.6352	0.8049	0.7608	5.0186	0.2664	10.9245

which results in

$$u_{ij} = \pi_{ij} \exp\left(-\frac{\alpha}{\lambda}(d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir})\right) \exp\left(-1 - \frac{\gamma_j}{\lambda}\right), \quad (\text{A.4})$$

where  $\alpha = (1 + \sum_{r \in N_j} \frac{1}{d_{jr}+1})$  is a constant for a fixed neighboring window. Utilizing  $\sum_{i=1}^c u_{ij} = 1$  to eliminate intermediate variable of  $\exp(-1 - \frac{\gamma_j}{\lambda})$ , we can obtain the membership at the  $(b+1)$ th step

$$u_{ij} = \frac{\pi_{ij} \exp\left(-\frac{\alpha}{\lambda}(d_{ij} + \sum_{r \in N_j} \frac{1}{N_R} d_{ir})\right)}{\sum_{k=1}^c \pi_{kj} \exp\left(-\frac{\alpha}{\lambda}(d_{kj} + \sum_{r \in N_j} \frac{1}{N_R} d_{kr})\right)}. \quad (\text{A.5})$$

#### APPENDIX B

##### PROOF OF CLUSTER CENTERS $\mu_i$

To estimate the clustering centers at  $(b+1)$ th step, we create the temporary function of  $J_{\mu_i}$  with respect to  $\mu_i$  according to (19) (ignoring constant terms)

$$J_{\mu_i} = \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} (x_r - \mu_i)^T \Sigma_i^{-1} (x_r - \mu_i)), \quad (\text{B.1})$$

In a similar fashion, we can get

$$\frac{\partial J_{\mu_i}}{\partial \mu_i} = \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} (x_j - \mu_i) + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} (x_r - \mu_i)) 2\Sigma_i^{-1} = 0, \quad (\text{B.2})$$

which results in

$$\begin{aligned} & \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} x_j + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} x_r) 2\Sigma_i^{-1} \\ &= 2\Sigma_i^{-1} \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)}) \mu_i, \end{aligned} \quad (\text{B.3})$$

Since there are  $2\Sigma_i^{-1}$  on both sides of the formula (B.3), we can delete the influence of  $2\Sigma_i^{-1}$ . Then, we get clustering centers at the  $(b+1)$ th step

$$\mu_i = \frac{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} x_j + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} x_r)}{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)})}. \quad (\text{B.4})$$

#### APPENDIX C

##### PROOF OF COVARIANCE MATRIX $\Sigma_i$

To estimate the covariance matrix at  $(b+1)$ th step, we create the temporary function of  $J_{\Sigma_i}$  with respect to  $\Sigma_i$  according to (19) (ignoring constant terms)

$$\begin{aligned} J_{\Sigma_i} &= \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (\log |\Sigma_i| \\ &+ t_{ij}^{(b)} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) \\ &+ \sum_{r \in N_j} \frac{1}{N_R} (\log |\Sigma_i| + t_{ir}^{(b)} (x_r - \mu_i)^T \Sigma_i^{-1} (x_r - \mu_i))), \end{aligned} \quad (\text{C.1})$$

Similarly, we can get

$$\begin{aligned} \frac{J_{\Sigma_i}}{\Sigma_i} &= \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (-t_{ij}^{(b)} (x_j - \mu_i)^T \Sigma_i^{-1} (x_j - \mu_i) \\ &- \mu_i) + 1 + \sum_{r \in N_j} \frac{1}{N_R} (-t_{ir}^{(b)} (x_r - \mu_i)^T \Sigma_i^{-1} (x_r - \mu_i) \\ &+ 1)) = 0, \end{aligned} \quad (\text{C.2})$$

which results in

$$\begin{aligned} & \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} (x_j - \mu_i)^T (x_j - \mu_i) \\ &+ \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} (x_r - \mu_i)^T (x_r - \mu_i)) \Sigma_i^{-1} = \sum_{j=1}^n (u_{ij} \\ &+ \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (1 + \sum_{r \in N_j} \frac{1}{N_R}) = 0, \end{aligned} \quad (\text{C.3})$$

Thus, the covariance matrix is expressed as

$$\Sigma_i = \frac{\sum_{j=1}^n \left[ (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (t_{ij}^{(b)} (x_j - \mu_i)^T (x_j - \mu_i) + \sum_{r \in N_j} \frac{1}{N_R} t_{ir}^{(b)} (x_r - \mu_i)^T (x_r - \mu_i)) \right]}{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (1 + \sum_{r \in N_j} \frac{1}{N_R})}, \quad (\text{C.4})$$

#### APPENDIX D

##### PROOF OF DEGREE OF FREEDOM $v_i$

To estimate the degree of freedom at  $(b+1)$ th step, we create the temporary function of  $J_{v_i}$  with respect to  $v_i$  according to (19) (ignoring constant terms)



$$J_{v_i} = \sum_{i=1}^c \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (\log \Gamma(\frac{v_i}{2}) - \frac{v_i}{2} \log(\frac{v_i}{2}) - (\frac{v_i+D}{2} - 1) (\log t_{ij}^{(b)} - \log(\frac{v_i^{(b)}+D}{2}) + \psi(\frac{v_i^{(b)}+D}{2}))) \\ + \frac{v_i}{2} t_{ij}^{(b)} + \sum_{r \in N_j} \frac{1}{N_R} (\log \Gamma(\frac{v_i}{2}) - \frac{v_i}{2} \log(\frac{v_i}{2}) + \frac{v_i}{2} t_{ij}^{(b)} - (\frac{v_i+D}{2} - 1) (\log t_{ir}^{(b)} - \log(\frac{v_i^{(b)}+D}{2}) + \psi(\frac{v_i^{(b)}+D}{2}))), \quad (D.1)$$

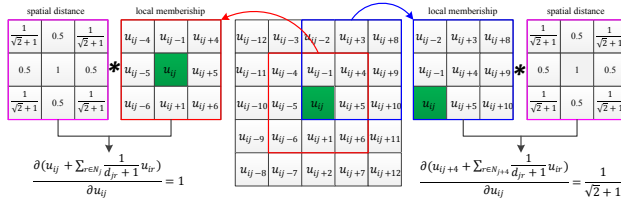


Fig. 10. Derivation of spatial local membership with  $3 \times 3$  window.

where  $v_i^{(b)}$  and  $t_{ij}^{(b)}$  are constants at  $(b+1)th$  step, since

$$\frac{\partial \log \Gamma(\frac{v_i}{2})}{\partial v_i} = \frac{1}{2} \psi(\frac{v_i}{2}), \quad (D.2)$$

$$\frac{\partial \frac{v_i}{2} \log(\frac{v_i}{2})}{\partial v_i} = \frac{1}{2} \log(\frac{v_i}{2}) + \frac{1}{2}, \quad (D.3)$$

$$\frac{\partial \frac{v_i}{2} t_{ij}^{(b)}}{\partial v_i} = \frac{v_i}{2} t_{ij}^{(b)}, \quad (D.4)$$

$$\frac{\partial (\frac{v_i+D}{2} - 1) (\log t_{ij}^{(b)} - \log(\frac{v_i^{(b)}+D}{2}) + \psi(\frac{v_i^{(b)}+D}{2}))}{\partial v_i} \\ = \frac{1}{2} (\log t_{ir}^{(b)} - \log(\frac{v_i^{(b)}+D}{2}) + \psi(\frac{v_i^{(b)}+D}{2})), \quad (D.5)$$

It is easy to obtain (D.6). Solving the formula, we can obtain

$$\psi(\frac{v_i}{2}) + \frac{\sum_{j=1}^n \left[ (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) ((t_{ij}^{(b)} - \log t_{ij}^{(b)}) + \sum_{r \in N_j} \frac{1}{N_R} (t_{ir}^{(b)} - \log t_{ir}^{(b)})) \right]}{\sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr}+1} u_{ir}) (1 + \sum_{r \in N_j} \frac{1}{N_R})} \\ - \log(\frac{v_i}{2}) - 1 + \log(\frac{v_i^{(b)}+D}{2}) - \psi(\frac{v_i^{(b)}+D}{2}) = 0. \quad (D.7)$$

## REFERENCES

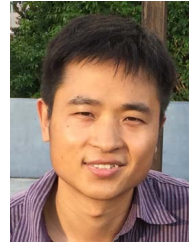
- [1] Y. Xu, E. Carlinet, T. Géraud, and L. Najman, "Hierarchical segmentation using tree-based shape spaces," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 39, no. 3, pp. 457-469, Mar. 2017.
- [2] K. Zhan, F. Nie, J. Wang, and Y. Yang, "Multiview consensus graph clustering," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1261-1270, Mar. 2019.
- [3] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972-976, Oct. 2007.
- [4] C. Xu, Z. Guan, W. Zhao, Y. Niu, Q. Wang, and Z. Wang, "Deep multi-view concept learning," in *Proc. Int. Joint Conf. Art. Intell. (IJCAI)*, Stockholm, Sweden, pp. 2898-2904, 2018.

- [5] X. Jia, T. Lei, X. Du, S. Liu, H. Meng, and A. K. Nandi, "Robust self-sparse fuzzy clustering for image segmentation," *IEEE Access*, vol. 8, pp. 146182-146195, 2020.
- [6] F. Zhao, Z. Zeng, H. Liu, R. Lan, and J. Fan, "Semisupervised approach to surrogate-assisted multiobjective kernel intuitionistic fuzzy clustering algorithm for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 6, pp. 1023-1034, Jun. 2020.
- [7] S. Zhou, D. Li, Z. Zhang, and R. Ping, "A new membership scaling fuzzy c-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, 2020, doi: 10.1109/TFUZZ.2020.3003441.
- [8] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, and W. Gao, "Multiple kernel k-means with incomplete kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1191-1204, May 2020.
- [9] J. Han, J. Xu, F. Nie, and X. Li, "Multi-view k-means clustering with adaptive sparse memberships and weight allocation," *IEEE Tran. Knowl. Data Eng.*, 2020, doi: 10.1109/TKDE.2020.2986201.
- [10] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, "A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data," *IEEE Trans. Med. Imag.*, vol. 21, no. 3, pp. 193-199, Mar. 2002.
- [11] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst. Man, Cybern., B, Cybern.*, vol. 34, no. 4, pp. 1907-1916, Aug. 2004.
- [12] S. Krinidis and V. Chatzis, "A robust fuzzy local information c-means clustering algorithm," *IEEE Trans. Image Process.*, vol. 19, no. 5, pp. 1328-1337, May 2010.
- [13] Y. Zhang, X. Bai, R. Fan, and Z. Wang, "Deviation-sparse fuzzy c-means with neighbor information constraint," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 1, pp. 185-199, Jan. 2019.
- [14] P. K. Mishro, S. Agrawal, R. Panda, and A. Abraham, "A novel type-2 fuzzy c-means clustering for brain MR image segmentation," *IEEE Trans. Cybern.*, 2020, doi: 10.1109/TCYB.2020.2994235.
- [15] M. Gong, Z. Zhou, and J. Ma, "Change detection in synthetic aperture radar images based on image fusion and fuzzy clustering," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2141-2151, Apr. 2011.
- [16] M. Gong, Y. Liang, J. Shi, W. Ma, and J. Ma, "Fuzzy c-means clustering with local information and kernel metric for image segmentation," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 573-584, Feb. 2013.
- [17] Z. Zhao, L. Cheng, and G. Cheng, "Neighbourhood weighted fuzzy c-means clustering algorithm for image segmentation," *IET Image Process.*, vol. 8, no. 3, pp. 150-161, Mar. 2014.
- [18] H. Zhang, Q. Wang, W. Shi, and M. Hao, "A novel adaptive fuzzy local information c-means clustering algorithm for remotely sensed imagery classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 55, no. 9, pp. 5057-5068, Jul. 2017.
- [19] Y. Tang, F. Ren, and W. Pedrycz, "Fuzzy c-means clustering through SSIM and patch for image segmentation," *Appl Soft Comput.*, vol. 87, pp. 105928, 2020.
- [20] L. Szilagyi, Z. Benyo, S. M. Szilagyi, and H. S. Adam, "MR brain image segmentation using an enhanced fuzzy c-means algorithm," in *Proc. 25th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pp. 17-21, 2003.
- [21] W. Cai, S. Chen, and D. Zhang, "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation," *Pattern Recognit.*, vol. 40, no. 3, pp. 825-838, Mar. 2007.
- [22] F. Zhao, J. Fan, and H. Liu, "Optimal-selection-based suppressed fuzzy c-means clustering algorithm with self-tuning non local spatial information for image segmentation," *Expert Sys. Appl.*, vol. 41, no. 9, pp. 4083-4093, Jul. 2014.
- [23] F. Guo, X. Wang, and J. Shen, "Adaptive fuzzy c-means algorithm based on local noise detecting for image segmentation," *IET Image Process.*, vol. 10, no. 4, pp. 272-279, 2016.



$$\begin{aligned} \frac{\partial J_{v_i}}{\partial v_i} = & \sum_{j=1}^n (u_{ij} + \sum_{r \in N_j} \frac{1}{d_{jr} + 1} u_{ir}) (\frac{1}{2} \psi(\frac{v_i}{2}) - (\frac{v_i}{2} \log(\frac{v_i}{2}) + \frac{1}{2}) + \frac{v_i}{2} t_{ij}^{(b)} - \frac{1}{2} (\log t_{ij}^{(b)} - \log(\frac{v_i^{(b)} + D}{2}) + \psi(\frac{v_i^{(b)} + D}{2})) \\ & + \sum_{r \in N_j} \frac{1}{N_R} (\frac{1}{2} \psi(\frac{v_i}{2}) - (\frac{v_i}{2} \log(\frac{v_i}{2}) + \frac{1}{2}) + \frac{v_i}{2} t_{ij}^{(b)} - \frac{1}{2} (\log t_{ij}^{(b)} - \log(\frac{v_i^{(b)} + D}{2}) + \psi(\frac{v_i^{(b)} + D}{2}))), \end{aligned} \quad (D.6)$$

- [24] T. Lei, X. Jia, Y. Zhang, L. He, H. Meng, and A. K. Nandi, "Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 3027-3041, Oct. 2018.
- [25] A. Abu and R. Diamant, "Enhanced fuzzy-based local information algorithm for sonar image segmentation," *IEEE Tran. Image Process.*, vol. 29, pp. 445-460, 2020.
- [26] K. S. Chuang, H. L. Tzeng, and S. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Comput Med. Imaging Graph.*, vol. 30, no. 1, pp. 9-15, Jan. 2013.
- [27] S. K. Adhikari, J. K. Sing, D. K. Basu, and M. Nasipuri, "Conditional spatial fuzzy C-means clustering algorithm for segmentation of MRI images," *Appl Soft Comput.*, vol. 34, pp. 758-769, Jun. 2015.
- [28] Y. Guo and A. Sengur, "NCM: Neutrosophic c-means clustering algorithm," *Pattern Recognit.*, vol. 48, no. 8, pp. 2710-2724, Aug. 2015.
- [29] C. Wang, W. Pedrycz, M. Zhou, and Z. Li, "Sparse regularization-based fuzzy c-means clustering incorporating morphological grayscale reconstruction and wavelet frames," *IEEE Trans. Fuzzy Syst.*, 2020, doi: 10.1109/TFUZZ.2020.2985930.
- [30] S. Miyamoto and M. Mukaidono, "Fuzzy c-means as a regularization and maximum entropy approach," in *Proc. 7th Int. Fuzzy Syst. Association World Congress.*, pp. 86-92, Jun. 1997.
- [31] S. P. Chatzis and T. A. Varvarigou, "A fuzzy clustering approach toward hidden markov random field models for enhanced spatially constrained image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 5, pp. 1351-1361, Oct. 2008.
- [32] H. Zhang, Q. M. J. Wu, Y. Zheng, T. M. Nguyen, and D. Wang, "Effective fuzzy clustering algorithm with Bayesian model and mean template for image segmentation," *IET Image Process.*, vol. 8, no. 10, pp. 571-581, Oct. 2014.
- [33] M. Gong, L. Su, M. Jia, and W. Chen, "Fuzzy clustering with a modified MRF energy function for change detection in synthetic aperture radar images," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 1, pp. 98-109, Feb. 2014.
- [34] G. Liu, Y. Zhang, and A. Wang, "Incorporating adaptive local information into fuzzy clustering for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3990-4000, Nov. 2015.
- [35] C. Wang, W. Pedrycz, Z. Li, and M. Zhou, "Kullback-Leibler divergence-based fuzzy c-means clustering incorporating morphological reconstruction and wavelet frames for image segmentation," *arXiv:2002.09479*, 2020.
- [36] T. Lei, X. Jia, Y. Zhang, S. Liu, H. Meng, and A. K. Nandi, "Superpixel-based fast fuzzy c-means clustering for color image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 9, pp. 1753-1766, Sept. 2019.
- [37] J. Gu, L. Jiao, S. Yang, and F. Liu, "Fuzzy double c-means clustering based on sparse self-representation," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 612-626, Apr. 2018.
- [38] T. Lei, P. Liu, X. Jia, X. Zhang, H. Meng, and A. K. Nandi, "Automatic fuzzy clustering framework for image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 9, pp. 2078-2092, Sept. 2020.
- [39] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492-1496, Jun. 2014.
- [40] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32-57, Sept. 1973.
- [41] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," Springer Science and Business Media, 2013.
- [42] T. Lei, X. Jia, T. Liu, S. Liu, H. Meng, and A. K. Nandi, "Adaptive morphological reconstruction for seeded image segmentation," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5510-5523, Nov. 2019.
- [43] T. M. Nguyen and Q. M. J. Wu, "A nonsymmetric mixture model for unsupervised image segmentation," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 751-765, Apr. 2013.
- [44] D. Peel and G. McLachlan, "Robust mixture modeling using the t distribution," *Stat. Comput.*, vol. 10, no. 4, pp. 339-348, Oct. 2000.
- [45] S. Shoham, "Robust clustering by deterministic agglomeration EM of mixtures of multivariate t-distributions," *Pattern Recognit.*, vol. 35, no. 5, pp. 1127-1142, May 2002.
- [46] H. Zhang, Q. M. J. Wu, T. M. Nguyen, and X. Sun, "Synthetic aperture radar image segmentation by modified student's t-mixture model," *IEEE Trans. Geosci. Remote Sensing.*, vol. 52, no. 7, pp. 4391-4403, Jul. 2014.
- [47] S. P. Chatzis, D. I. Kosmopoulos, and T. A. Varvarigou, "Robust sequential data modeling using an outlier tolerant hidden markov model," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 9, pp. 1657-1669, Sept. 2009.
- [48] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898-916, 2011.
- [49] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Graz, Austria, 2006, pp. 1-15.
- [50] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603-619, May 2002.



**Tao Lei** (M'17-SM'20) received the Ph.D. degree in Information and Communication Engineering from Northwestern Polytechnical University, Xi'an, China, in 2011. From 2012 to 2014, he was a Post-doctoral Research Fellow with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. From 2015 to 2016, he was a Visiting Scholar with the Quantum Computation and Intelligent Systems group at University of Technology Sydney, Sydney, Australia. He has authored and co-authored 80+ research papers including IEEE

TIP, TFS, TGRS, TGRSL, ICASSP, ICIP, and FG.

He is currently a Professor with the School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology. His current research interests include image processing, pattern recognition, and machine learning.



**Xiaohong Jia** received the M.S. degree in Signal and Information Processing from Lanzhou Jiaotong University, Lanzhou, China, in 2017. He is currently pursuing the Ph.D. degree at the School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China.

His current research interests include image processing and pattern recognition.



**Dinghua Xue** received the M.S. degree from Shaanxi University of Science and Technology, Xi'an, China. She is currently pursuing the Ph.D. degree at the School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an 710021, China. Her current research interests include image processing and pattern recognition.



**Qi Wang** (M'15-SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science, and with the Center for OPTical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.



**Hongying Meng** (M'10-SM'17) received his Ph.D. degree in Communication and Electronic Systems from Xi'an Jiaotong University, Xi'an China. He is an associate editor for IEEE Transactions on Circuits and Systems for Videos Technology (TCSVT) and IEEE Transactions on Cognitive and Developmental Systems (TCDS). He is a Fellow of The Higher Education Academy (HEA) and a member of Engineering Professors Council in UK. He has authored over 130 publications including IEEE TIP, TCYB, TFS, TAC, TCSVT, TBE, TCDS, ICASSP and CVPR.

He is currently a Reader at the Department of Electronic and Electrical Engineering, Brunel University London, U.K. His research interests include digital signal processing, affective computing, machine learning, human computer interaction, and computer vision.



**Asoke K. Nandi** (F'11) received the degree of Ph.D. in Physics from the University of Cambridge (Trinity College), Cambridge (UK). He held academic positions in several universities, including Oxford (UK), Imperial College London (UK), Strathclyde (UK), and Liverpool (UK) as well as Finland Distinguished Professorship in Jyväskylä (Finland). In 2013, he moved to Brunel University London (UK), to become the Chair and Head of Electronic and Computer Engineering. Professor Nandi is a Distinguished Visiting Professor at Tongji University (China) and an Adjunct Professor at University of Calgary (Canada).

In 1983 Professor Nandi co-discovered the three fundamental particles known as  $W^+$ ,  $W^-$  and  $Z^0$  (by the UA1 team at CERN), providing the evidence for the unification of the electromagnetic and weak forces, for which the Nobel Committee for Physics in 1984 awarded the prize to his two team leaders for their decisive contributions. His current research interests lie in signal processing and machine learning, with applications to communications, image segmentations, biomedical data, etc. He has made many fundamental theoretical and algorithmic contributions to many aspects of signal processing and machine learning. He has much expertise in "Big and Heterogeneous Data", dealing with modelling, classification, estimation, and prediction.

He has authored over 600 technical publications, including 250 journal papers as well as five books, entitled Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines (Wiley, 2020), Automatic Modulation Classification: Principles, Algorithms and Applications (Wiley, 2015), Integrative Cluster Analysis in Bioinformatics (Wiley, 2015), Blind Estimation Using Higher-Order Statistics (Springer, 1999), and Automatic Modulation Recognition of Communications Signals (Springer, 1996). The H-index of his publications is 80 (Google Scholar) and his ERDOS number is 2.

Professor Nandi is a Fellow of the Royal Academy of Engineering (UK) as well as a Fellow of seven other institutions, including the IEEE and the IET. Among the many awards he received are the Institute of Electrical and Electronics Engineers (USA) Heinrich Hertz Award in 2012, the Glory of Bengal Award for his outstanding achievements in scientific research in 2010, the Water Arbitration Prize of the Institution of Mechanical Engineers (UK) in 1999, and the Mountbatten Premium, Division Award of the Electronics and Communications Division, of the Institution of Electrical Engineers (UK) in 1998. Professor Nandi is an IEEE EMBS Distinguished Lecturer (2018-19).